A Concept-Aware API for the Taxonomic Checklist Infrastructure in Germany

Elisabeth Hüllbusch[‡], Anton Güntsch[§], Katja Luther[§], Rudolf May[‡], Andreas Müller[§], Falko Riemenschneider[|], Gerhard Ludwig[‡]

‡ German Federal Agency for Nature Conservation, Bonn, Germany
§ Freie Universität Berlin, Botanic Garden and Botanical Museum Berlin, Berlin, Germany
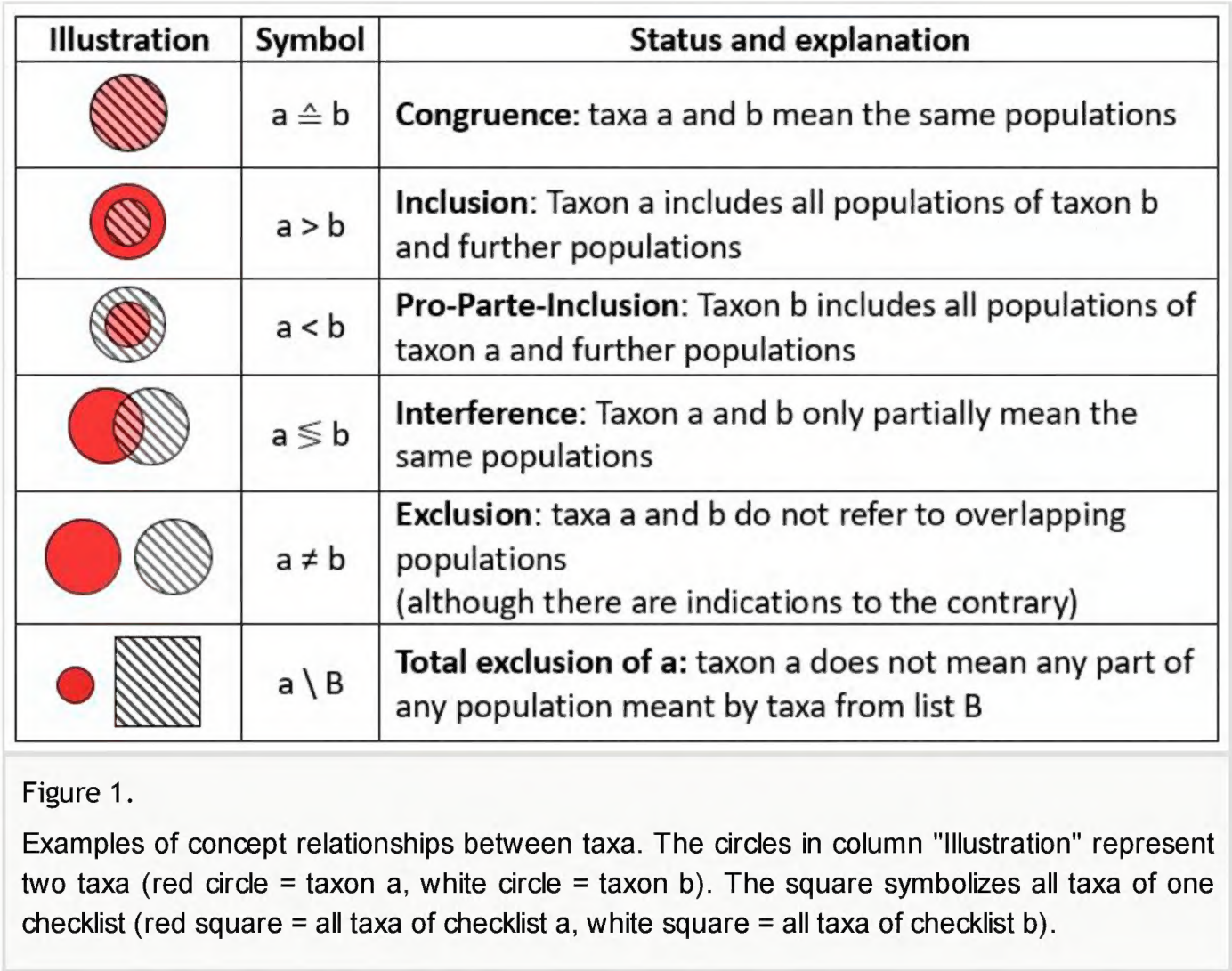| doctronic GmbH & Co. KG, Bonn, Germany

Corresponding author: Elisabeth Hüllbusch (elisabeth.huellbusch@bfn.de),
Anton Güntsch (a.guentsch@bo.berlin)

Abstract

Biological data refer to taxa (e.g., species) or to populations (e.g., all individuals of a taxon in a region). Scientific names of organisms are commonly used as identifiers for those data, e.g., in databases. However, scientific names are assigned (linked) to taxonomic concepts, which very often change over time or space, e.g., when a species name refers to different infraspecific taxa in different regions (Geoffroy and Berendsohn 2003). For this reason, the names themselves are not sufficient to unambiguously assign factual data to the taxonomic concepts they belong to. The problem has been discussed for many years, and data models such as the Berlin Taxonomic Information Model (Berendsohn et al. 2003) and the Common Data Model of the EDIT Platform for Cybertaxonomy (Güntsch et al. 2018) allow taxonomic concepts to be represented and linked. Taxon concepts are considered here as sets of (potential) specimens. Thus, relationships between taxa can be handled the same as mathematical set relationships (Fig. 1).

Illustration	Symbol	Status and explanation
	$a \triangleq b$	**Congruence**: taxa a and b mean the same populations
	$a > b$	**Inclusion**: Taxon a includes all populations of taxon b and further populations
	$a < b$	**Pro-Parte-Inclusion**: Taxon b includes all populations of taxon a and further populations
	$a \lesseqgtr b$	**Interference**: Taxon a and b only partially mean the same populations
	$a \neq b$	**Exclusion**: taxa a and b do not refer to overlapping populations (although there are indications to the contrary)
	$a \setminus B$	**Total exclusion of a:** taxon a does not mean any part of any population meant by taxa from list B

Figure 1.

Examples of concept relationships between taxa. The circles in column "Illustration" represent two taxa (red circle = taxon a, white circle = taxon b). The square symbolizes all taxa of one checklist (red square = all taxa of checklist a, white square = all taxa of checklist b).

From a user perspective, however, there are hardly any interfaces that can be used to effectively access taxonomic concepts and their relationships, for example to semantically annotate research data with name references. The German Federal Agency for Nature Conservation (BfN) and the National Research Data Infrastructure NFDI4Biodiversity have cooperatively developed an API (Application Programming Interface) for accessing available German species checklists, where taxonomic concept relationships between checklist versions can flexibly be queried and integrated into other systems. The query for scientific names includes a similarity search and initially returns the taxa with matching names. In a subsequent query, hierarchies, synonyms and the concept relationships to name usages in other checklist versions are output (e.g., congruence, inclusion, pro-parte-inclusion, interference, exclusion, and total exclusion).

From this output, it is possible to recognize, in particular, to which taxa an assignment of information is problematic.

The API is currently in the testing phase and is used in the context of NFDI4Biodiversity for the annotation and integration of research data.

Keywords

taxonomic concepts, taxa, taxon names, annotation

Presenting author

Elisabeth Hüllbusch

Presented at

SPNHC TDWG 2024

Funding program

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Berendsohn WG, Döring M, Geoffroy M, Glück K, Güntsch A, Hahn A, Kusber W-H, Li J-L, Röpert D, Specht F (2003) The Berlin Taxonomic Information Model. Schriftenreihe Vegetationskunde 39: 15-42. URL: https://www.researchgate.net/publication/281269078_The_Berlin_Model_a_concept-based_taxonomic_information_model
- Geoffroy M, Berendsohn WG (2003) The concept problem in taxonomy: importance, components, approaches. Schriftenreihe Vegetationskunde 39: 5-13. URL: https://www.researchgate.net/profile/Walter-Berendsohn/publication/285456910_The_concept_problem_in_taxonomy_Importance_components_approaches/links/5893193792851c5457479d9d/The-concept-problem-in-taxonomy-Importance-components-approaches.pdf
- Güntsch A, Müller A, Plitzner P, Henning T, Kilian N, Fichtmüller D, Gleisberg M, Karam N, Müller-Birn C, Berendsohn W (2018) The EDIT Platform for Cybertaxonomy - an integrated software environment for biodiversity research data management. In: Gaikad J, König-Ries B, Recknagel F (Eds) Proceedings 10th International Conference on Ecological Informatics. https://doi.org/10.22032/dbt.37806